A Theoretical Study on Bridging Internal Probability and Self-Consistency for LLM Reasoning

Zhi Zhou [1] **Yuhao Tan** [1] **Zenan Li** [2] **Yuan Yao** [1]
Lan-Zhe Guo [1] **Yu-Feng Li** ✉ [1] **Xiaoxing Ma** ✉ [1]

[1] State Key Laboratory of Novel Software Technology, Nanjing University, China
[2] Department of Computer Science, ETH Zurich, Switzerland

✉ Corresponding Author

The 39th Conference on Neural Information Processing Systems (NeurIPS 2025)

Background: LLM Reasoning

- LLMs typically **sample a reasoning path** and then **parse the answer** when reasoning.

Problem $\boldsymbol{x}$: Preliminary investigation revealed that during a public first aid operation, the defendant, Li, took advantage of the victim's inattention and used a razor blade to slash open the victim's clothing bag, stealing…

LLM Sampling $\hat{\boldsymbol{t}} \sim p(\boldsymbol{x}; \boldsymbol{\theta}_{\mathrm{LLM}})$

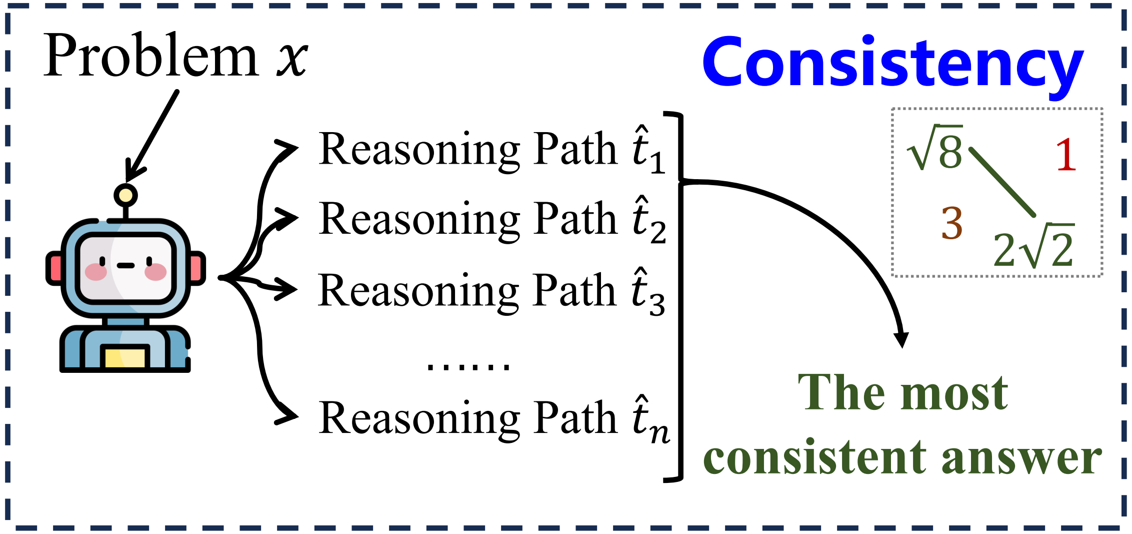
Reasoning path $\hat{\boldsymbol{t}}$: Based on the facts you provided, the amount involved in this case is calculated as follows: 1. Value of the stolen mobile phone: appraised at RMB 2,800; 2. Value of the stolen cash: RMB 500. Total amount involved = Mobile phone value + Cash value = RMB 2,800 + RMB 500 = RMB 3,300.

Reasoning answer $\hat{\boldsymbol{y}}$: 3300 RMB

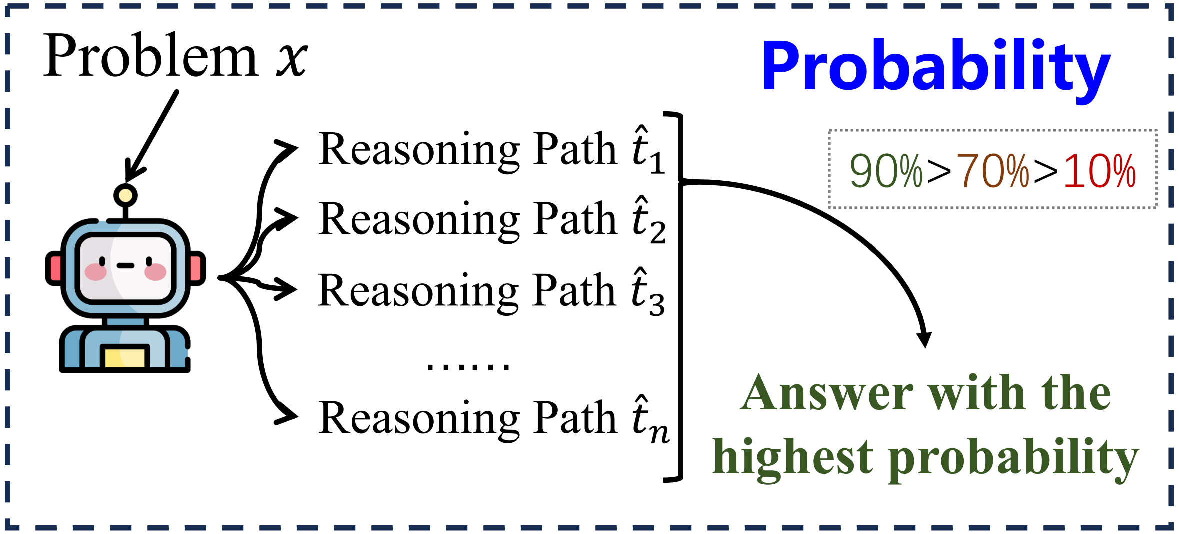
Parsing $\hat{\boldsymbol{y}} = g(\hat{\boldsymbol{t}})$

Background: LLM Reasoning

- **Sampling-based test-time scaling methods** further leverage multiple reasoning paths to enhance reasoning performance.
- Two typical sampling-based test-time scaling methods are:



Self-Consistency Method



Perplexity Method

Theoretical Framework

How can different test-time scaling methods for LLM reasoning be compared theoretically?

- **First theoretical framework for LLM reasoning** in context of confidence estimation:

  a) Treat both consistency and probability as confidence estimation $\hat{p}(\hat{y} \mid \boldsymbol{x})$;

  b) Compare them with the ground truth using mean squared error (MSE);

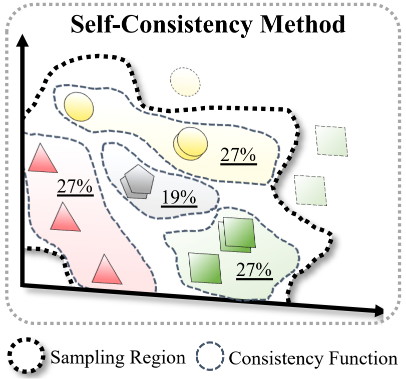
  $$\mathcal{E}_{\hat{p}}(\hat{y}) = \mathbb{E}\left[\left(\hat{p}(\hat{y} \mid x) - \mathbb{I}[\hat{y} = y]\right)^2\right]$$

  c) Decompose reasoning into **estimation error** (which is only related to the algorithm) and **model error** (which is only related to the model).

$$\mathcal{E}_{\hat{p}}(\hat{y}) = \underbrace{\mathbb{E}\left[\left(\hat{p}(\hat{y} \mid x) - p(\hat{y} \mid x)\right)^2\right]}_{\textit{Estimation Error}} + \underbrace{\left(p(\hat{y} \mid x) - \mathbb{I}[\hat{y} = y]\right)^2}_{\textit{Model Error}},$$

Theoretical Insights

- **Self-consistency methods (confidence is estimated based on consistency)**

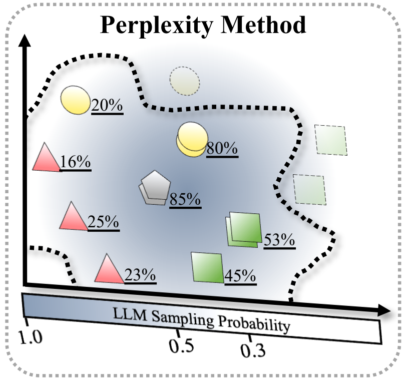


$$\mathcal{E}_{\hat{p}^{(\mathrm{SC})}}(\hat{y}) = \underbrace{\frac{1}{n} p(\hat{y} \mid x)(1 - p(\hat{y} \mid x))}_{\textit{Estimation Error}} + \underbrace{\left(p(\hat{y} \mid x) - \mathbb{I}[\hat{y} = y]\right)^2}_{\textit{Model Error}}$$

Estimation error: Linear decay

Model error: Consider the structure of the answer space through parsing function g

- **Perplexity-based methods (confidence is estimated based on probability)**



$$\mathcal{E}_{\hat{p}^{(\mathrm{PPL})}}(\hat{t}) = \underbrace{(1 - p(\hat{t} \mid x))^n p(\hat{t} \mid x)(2\mathbb{I}[\hat{y}_i = y] - p(\hat{t} \mid x))}_{\textit{Estimation Error}} + \underbrace{\left(p(\hat{t} \mid x) - \mathbb{I}[g(\hat{t}) = y]\right)^2}_{\textit{Model Error}}$$

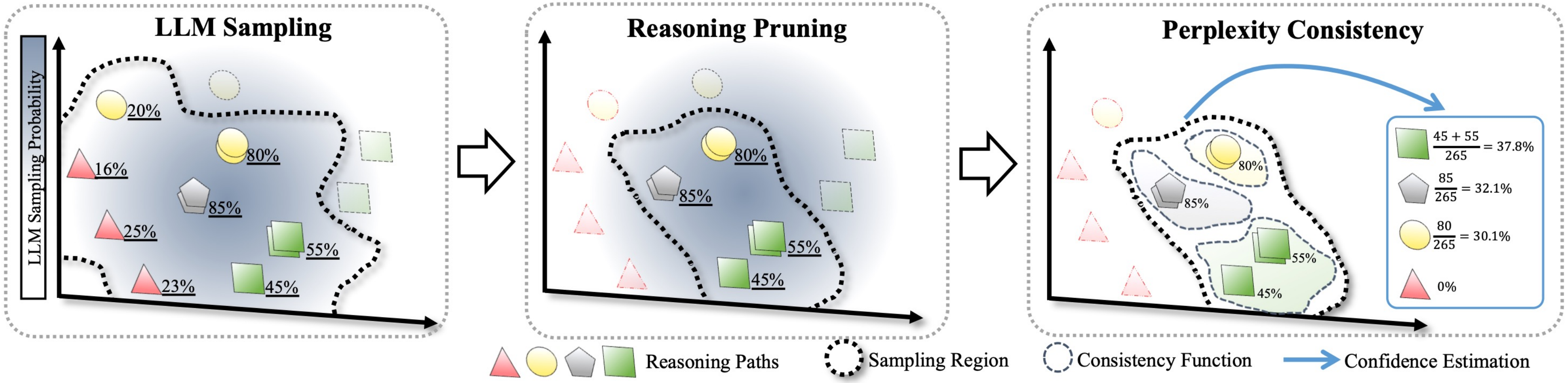
Estimation error: Exponential decay

Model error: Cannot utilize prior knowledge of parsing function g

Method

- **Reasoning Pruning Perplexity Consistency (RPC)**

—— a simple approach to combine advantages of self-consistency and perplexity



$$\mathcal{E}_{\hat{p}^{(\mathrm{PC})}}(\hat{y}) = \underbrace{\alpha^n p(\hat{y}\,|\,x)\big(2\mathbb{I}[\hat{y}=y] - (1+\alpha^n)p(\hat{y}\,|\,x)\big)}_{\textit{Estimation Error}} + \underbrace{\left(p(\hat{y}\,|\,x) - \mathbb{I}[\hat{y}=y]\right)^2}_{\textit{Model Error}}.$$

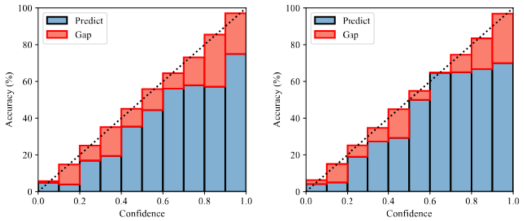
Estimation Error: Exponential decay; **Model Error**: Identical to that of self-consistency method.

Experiments

Table 1: Efficiency comparison of *Perplexity Consistency* module (PC) and RPC. The table shows the minimum number of samples needed to exceed the best performance of SC, with reduction rates in bold when sampling is reduced.

Method	MATH		MathOdyssey		OlympiadBench		AIME	
	Accuracy	#Samplings	Accuracy	#Samplings	Accuracy	#Samplings	Accuracy	#Samplings
Best of SC	50.57	64	28.32	112	11.07	128	9.40	128
PC	50.63	32	28.51	112	11.07	128	9.00	64
Δ	+0.06	**-50.0%**	+0.19	-0.0%	0.00	-0.0%	0.00	**-50.0%**
RPC	51.16	32	29.31	32	11.07	64	9.50	48
Δ	+0.59	**-50.0%**	+0.99	**-71.4%**	0.00	**-50.0%**	+0.10	**-62.5%**

Efficiency: Reduce 50% sampling overhead for LLM reasoning at the same performance level.

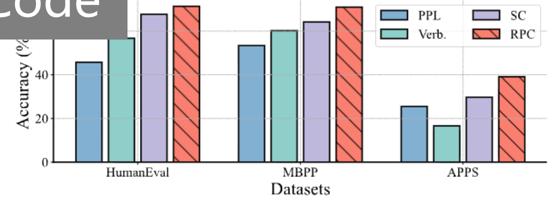


(a) SC (b) RPC

Figure 3: The reliability diagrams of SC and RPC on MathOdyssey dataset using InternLM-2-MATH-Plus 7B model.

Reliability: Improve the LLM reasoning reliability metric.

Code



Commonsense

		iQA
PPL	41.46	54.36
SC	43.00	56.71
RPC	**44.09**	**58.42**

R1 Model

	hOdyssey	AIME	MATH	OlympiadBench
PPL	60.04	72.92	81.81	21.65
SC	57.22	70.40	82.03	21.93
RPC	**61.11**	**76.47**	**82.78**	**22.81**

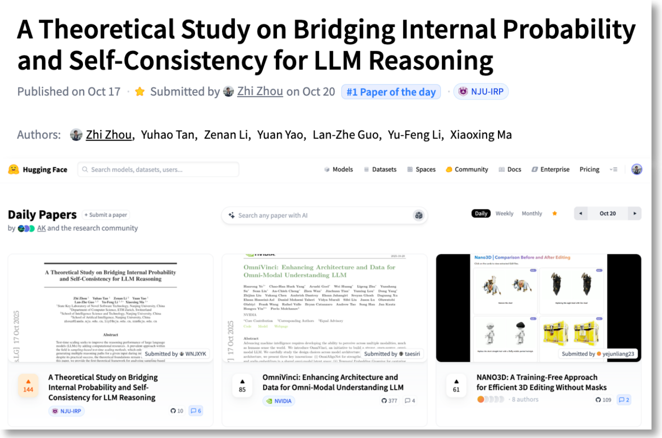
Generality: Effectiveness has been validated on code generation and commonsense reasoning tasks. Additionally, this approach is applicable to recent R1 LLMs.

Conclusion

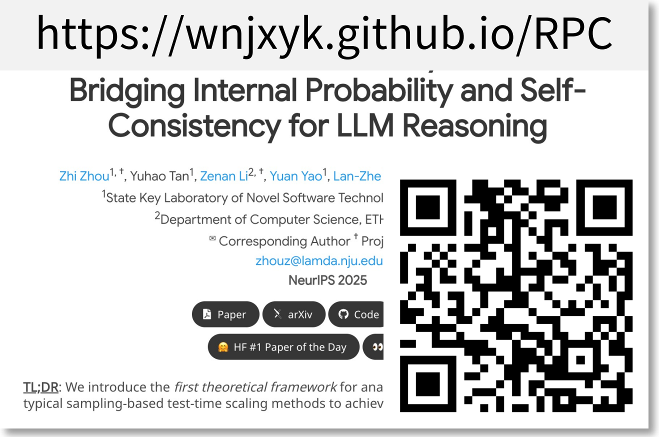
- This paper presents the **first theoretical framework for analyzing test-time scaling methods** in LLM reasoning.
- We propose a novel method that combines the advantages of two existing methods.
- Experimental results on mathematical, code generation, and commonsense reasoning tasks demonstrate the effectiveness of our method.

Thanks!

Hugging Face #1 Paper of The Day



Homepage



Demo